Lysine Directed Cross-Linking of Viral DNA−RNA:DNA Hybrid Substrate to the Isolated RNase H Domain of HIV-1 Reverse Transcriptase[†]

Juan P. Guaitiao,[‡] Roberto A. Zúñiga,[‡] Monica J. Roth,[§] and Oscar Leon*[,‡]

Programa de Virología, Instituto de Ciencias Biomédicas, Facultad de Medicina, Universidad de Chile, Independencia 1027, Santiago, Chile, and Department of Biochemistry, University of Medicine and Dentistry of New Jersey−Robert Wood Johnson Medical School, 675 Hoes Lane, Piscataway, New Jersey 08854 USA

Received August 14, 2003; Revised Manuscript Received November 21, 2003

ABSTRACT: An isolated ribonuclease H domain of HIV-1 reverse transcriptase is capable of specifically removing the tRNA primer within an oligonucleotide mimic. The determinants for substrate specificity are located in a region within the terminal octanucleotide of the acceptor stem of the tRNA. Recognition of the substrate by HIV-1 RNase H was analyzed by the introduction of a cross-linking reagent directed toward lysines on the thymine residue complementary to the scissile bond, facing the major groove of the DNA−RNA:DNA substrate. Cross-linking of the modified substrate to RNase H required the presence of Mn^{2+}. The Mn^{2+} titration of cross-linking paralleled the Mn^{2+} requirement for activity. Modified substrate quenched with glycine prior to binding of substrate was efficiently cleaved, whereas the RNA within the cross-linked product was intact. Tryptic digestion of the isolated RNase H−nucleic acid covalent complex revealed a main cross-linked peptide whose N-terminal peptide sequence is VVTLTDTTNQ, indicating that the cross-linked lysine corresponds to Lys476. Cross-linking to K476 was confirmed by analysis of K476C RNase H. Mutation of K476C disrupted the chemical cross-linking while maintaining activity. On the basis of the size of the cross-linker arm, the results indicate that K476 is in closer proximity to the tRNA mimic substrate within the isolated RNase H domain than observed for the RNase H-resistant polypurine tract (PPT) substrate within the HIV-1 RT.

Retroviral reverse transcriptases (RT)[1] catalyze the synthesis of a double-stranded DNA utilizing the single-stranded viral RNA genome as a template. This DNA intermediate is subsequently processed and integrated into the host genome. The general mechanism of reverse transcription involves two enzyme activities present in RT:RNA and DNA dependent DNA polymerase and RNase H. RNase H is required for removal of the viral RNA, primer formation and removal for the synthesis of the (+) strand DNA, and specific removal of the primer tRNA used in the synthesis of the (−) strand DNA (for a review see ref *1*).

HIV-1 RT is a heterodimer composed of 66 and 51 kDa subunits. The p66 subunit contains the RNase H domain at the C-terminus and contributes the polymerase activity. Cross-linking experiments suggest that p51 is involved in binding of the tRNA, relevant in the initiation of replication (*2*). The 3D structures of RT complexed to a dsDNA and an RNase H resistant RNA:DNA hybrid (HIV−PPT) have been reported (*3*, *4*). The contacts between RT and the DNA primer in these complexes are similar; however, there are some differences in the contacts with the RNA template. The nucleic acids in both cases have a bend of 40° (*4*). Studies on the crystal structure of PPT dsDNA and RNA:DNA duplexes (*5*, *6*) have revealed that A-tracts impart a B′ form to duplex DNA and DNA:RNA, yielding narrower minor grooves than the H-form seen in RNA:DNA hybrids without these A-tracts (*7*−*9*). A polypurine tract nucleic acid duplex (either as dsDNA or RNA:DNA hybrid), therefore, is a structurally specialized moiety and is not representative of generalized sequence duplexes. The structure of the RT−PPT RNA:DNA complex therefore, does not provide insights regarding specific contacts required for catalysis, as this hybrid is not cleaved by RNase H. This report provides a structural basis to understand the recognition of a cleavable RNA:DNA substrate.

Previous studies have shown that an isolated RNase H domain can catalyze the removal of tRNA sequences in a mimic DNA−RNA:DNA hybrid substrate, similar to the full-length RT in the absence of polymerization (*10*). Mutagenesis on the mimic substrate allowed us to identify the determinants on the RNA primer for specific recognition by RNase H. The results indicated that the region corresponding to the terminal nine ribonucleotides of the acceptor stem contains the recognition elements (*11*). Within this region, positions +4 and +6 were important for the specificity of cleavage by HIV-1 RNase H.

As a first approach to determine the regions of RNase H that may be in contact with the DNA−RNA:DNA hybrid next to the DNA−RNA junction, we carried out cross-linking

[†] This work was supported by Grants RO1 CA90174 Award and Fondecyt 198982.

* Corresponding author. Tel: 56 63 2 678 6858. Fax 56 63 2 678 6124. E-mail: oleon@med.uchile.cl.

[‡] Universidad de Chile.

[§] University of Medicine and Dentistry of New Jersey−Robert Wood Johnson Medical School.

[1] Abbreviations: C2dT, 5′-dimethoxytrityl-5-[*N*-(trifluoroacetyl-aminoethyl)-3-acrylimido]-2′-deoxyuridine, 3′-[(2-cyanoethyl)-(*N*,*N*-diisopropyl)]-phosphoramidite; DSG, disuccinimidyl glutarate HIV-1, human immunodeficiency virus type 1; m1A, 1-methyl adenosine; PPT, polypurine tract; RT, reverse transcriptase; RNase H, ribonuclease H.

10.1021/bi035454y CCC: $27.50
Published on Web 01/13/2004

studies. Here we present the results of a lysine directed cross-linking utilizing a bifunctional reagent attached to a base located in the DNA strand at the cleavage site (position +1). The cross-linked peptide was identified by amino acid sequencing and encoding a single Lys available for cross-linking, Lys476. This peptide also contains Glu 478 and Glu 498 that belong to the putative active site of RNase H. This result indicates that the positioning of the DNA−RNA:DNA hybrid in the complex within the isolated RNase H domain differs from that found in the PPT RNA:DNA−RT complex. A model for substrate binding is discussed.

EXPERIMENTAL PROCEDURES

Chemicals and Enzymes. Disuccinimidyl glutarate (DSG) was purchased from Pierce (Rockford, IL), and Centricon 30 and nitrocellulose membranes were from Millipore (MA). DEAE membranes were from Sartorious Corp. (New York, NY). [γ^{32}P]-ATP was purchased from NEN (MA). Polynucleotide kinase (PNK) was purchased from New England Biolabs (MA), trypsin sequencing grade and RNasin was from Promega (Madison, WI). NY 427 (isolated RNase H domain) was purified from *Escherichia coli* BL21-DE3 harboring the plasmid pETNY427 (*10*). K476C RNase H was purified from the soluble fraction as previously described for NY427 (*10*) and eluted from NTA nickel affinity column (Qiagen) in the presence of 300 mM imidazole. Thrombin was from Sigma Chemical Co. (St. Louis, MO).

Substrate Oligonucleotides. 5′ AGCAGXGGCGCCCG 3′ where X corresponds to the modified base amino-modifier C2dT (Glen Research) was synthesized in Ana-Gen Tech (Palo Alto, CA). The RNA−DNA oligonucleotide 5′ **CGGGCGCCA**CTGCT 3′ was synthesized at Cybersyn (Lenni, PA). 5′ AGCAGTGGCGCCCG 3′ was synthesized by the DNA synthesis facility, Department of Biochemistry, UMDNJ. Oligos 5′GGTACCAGCATGCAAAGGAAT 3′ and 5′ ATTCCTTTGCATGCTGGTACC 3′ (21-mers) were obtained from Invitrogen.

RNase H Cleavage Assays. Preparation and labeling of the hybrid DNA−RNA:DNA substrate mimic was done as previously described (*11*). A total of 1 pmol of the annealed hybrid was incubated at 37 °C with 1 pmol of NY 427 in 15 μL of a buffer containing 50 mM Hepes pH 7.8, 2 mM DTT, 8 mM $MnCl_2$. Aliquots (2 μL) were added to 5 μL of formamide stop buffer and the products were analyzed by electrophoresis in 20% acrylamide−7 M urea gels.

Preparation of the Cross-Linking Derivative. For cross-linking, the DNA oligonucleotide containing the base C2dT was 5′ labeled with [γ^{32} P]-ATP. The excess of ATP was removed by filtration on Sephadex G-25 and the oligonucleotide was annealed to the RNA−DNA strand in 50 mM Hepes pH 7.8, 50 mM KCl, 8 mM $MgCl_2$. DSG coupling to C2dT was carried out as decribed (*12*). The reaction contained 50 mM Hepes pH 7.8, 40% DMSO, 1−8 μM DNA−RNA:DNA substrate and 20 mM DSG. DSG was dissolved in DMSO and was the last component added. After 10 min at 25 °C, the modified substrate was precipitated by adding 0.3 vol of 4 M sodium acetate, pH 6.0 and 3 vol of ethanol, chilled at −70 °C for 20 min and centrifuged at 14000*g* for 30 min at 4 °C. The supernatant was discarded and the pellet was washed twice with 1 mL of 80% ethanol. The DSG modified oligonucleotide was dissolved in nuclease-free water just prior to use. When indicated, the DSG modified substrate was incubated with 50 mM glycine, prior to ethanol precipitation.

Cross-Linking. Cross-linking was carried out in Hepes pH 7.8, 8 mM $MnCl_2$ at 25 °C (1 pmol of NY427 and 1 pmol of the nucleic acid derivative in 20 μL). At desired times, aliquots were taken for analysis of the cross-linking products by electrophoresis in polyacrylamide gels with SDS or by binding to nitrocellulose filters (Millipore type HA) in the presence of high salt as described (*13*).

Thrombin Digestion. Seven microliters of the cross-linking reaction (see above) were added to 13 μL of 40 mM Hepes pH 7.8, 20 mM glutamate, 0.2% NP-40, 3 mM DTT, 2 mM EDTA, 1 mM $CaCl_2$, 40% glycerol. A total of 0.1 U thrombin was added and incubated at 25 °C for 1 h. The reaction was stopped by the addition of 20 μL of 0.12 M Tris HCl, pH 6.8, 20% glycerol, 0.2 M DTT, 0.005% bromophenol blue, and 40 mM EDTA.

Purification of the Cross-Linked Peptides. Large-scale reactions (1.15 mL) for peptide purification involved cross-linking 4 nmol of NY427 with 4 nmol of DSG modified oligonucleotide (15 cpm/pmol) for 30 min at 25 °C. The reaction was quenched with 50 mM glycine and the sample was concentrated in Centricon 30 tubes, washed with 4 M urea, and loaded (80 μL) directly onto a 12% acrylamide−7 M urea gel. The RNase H−DNA complex was visualized by autoradiography, and the corresponding band was excised and eluted in 0.8 mL of water for 16 h at 37 °C. A total of 50 μg of NY427 were added to the eluate as a carrier and after concentration in a Centricon 30 tube, the complex was washed with 0.1 M NH_4HCO_3 to a final volume of 150 μL. Finally, 50 μL of 8 M urea and 3% w/w trypsin were added. This mixture was incubated at 37 °C for 16 h, and filtrated on a Q5F anion exchanger (Sartorius) equilibrated with 10 mL of K_2HPO_4 pH 6.6, 4.8% ethanol, 40 mM KCl. The membrane was washed sequentially with 1.5 mL of the equilibration solution, 5 mL of 20 mM K_2HPO_4 pH 6.6, 4.8% ethanol, 0.2 M KCl. Finally, the membrane was washed with 2 mL of 20 mM K_2HPO_4 pH 6.6, 4.8% ethanol, 0.8 M KCl. This fraction, that contained most of the radioactivity, was passed through a C18 cartridge (Waters) equilibrated with 20 mM TEAA, pH 6.85 (20 mM triethylamine-acetate). The cartridge was washed with 3 mL of the same solution and eluted with 1 mL of 50% ethanol. This fraction was evaporated to dryness, dissolved in 50 μL of 20 mM TEAA pH 6.85, passed through a Sephadex G-25 spin column (Boehringer) equilibrated with 20 mM TEAA pH 6.85, and evaporated to dryness. This sample was subjected to amino acid sequencing at the W. M. Keck Facility, Yale University.

Isolation of RNA from the Cross-Linked Complex. Cross-linking reactions were assembled using a 5′-end labeled RNA strand hybridized to the C2dT modified DNA. The cross-linked complex was isolated under native conditions essentially as described (*14*) except that SDS and MCE were not added and the samples were not heated prior to loading onto a 15% PAG. The complex was excised from the gel in DEPC-treated H_2O and heated at 70 °C for 15 min, followed by overnight elution at room temperature. Samples were loaded onto a 20% PAG containing 7 M urea.

Mutagenesis of K476C. Plasmid pET-NY427 (*10*) was used as a template for overlapping PCR utilizing mutagenic oligonucleotide K476-1 (5′ CAACAAATCAGTGTACTGAG-

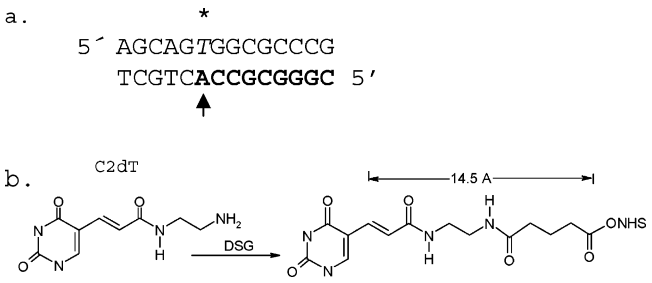

FIGURE 1: Scheme of the modifications used for attachment of DSG. (a) Model substrate for tRNA removal. The T in italics (*) indicates the position of the modified base C_2dT. The arrow marks the position of RNase H cleavage. (b) Coupling of DSG to the amino group of C_2dT.

TTA 3′) and K476-2 (5′ TAACTCAGTACACTGATTTGTTG 3′) and outside T7 primers T7-term (5′ CTAGTTATTGCTCAGCGGTG 3′) and T7-prom (5′ CCCGCGAAATTAATACGACT 3′). Mutations were exchanged into pET-NY427 using the NdeI−BamHI fragment. Mutations were confirmed by DNA sequence analysis. Oligonucleotides to generate K476C were synthesized by Bioschile.

RESULTS

Reverse transcription of the HIV-1 genome by the reverse transcriptase requires the specific recognition and removal of DNA−RNA:DNA hybrids by the RNase H domain of RT. The positioning and recognition of these specific sequences within the protein is a key area of investigation. Minimally, recognition and removal of the tRNA:DNA intermediate are maintained using the isolated HIV-1 RNase H domain. To further define this recognition, cross-linking of an tRNA:DNA mimic with the HIV-1 RNase H domain was initiated.

Synthesis of the Mimic DNA−RNA:DNA Hybrid and Derivatization. The substrate mimic for the tRNA removal containing the modified base C2dT is outlined in Figure 1A. The substrate consists of the 3′ nine terminal RNA sequences of the $tRNA^{Lys,3}$ covalently linked to five deoxynucleotides encoded by the U5 LTR hybridized with a complementary DNA oligonucleotide. Specific RNase H cleavage of the substrate occurs between the C and A residue at the 3′ terminus of the RNA (Figure 1A, arrow). The modified base is within the DNA, complementary to the A at the scissile bond (Figure 1A, star). Figure 1B shows the structure and the length of the cross-linker within the mimic DNA−RNA:DNA hybrid. C2dT replaces the 5′ methyl group with a side chain containing an amino group, facilitating the coupling of DSG. The modified base would react to lysine residues within 14.5 Å of the bound substrate. The side chain of C2dT faces the major groove in the AT base pair of the DNA−RNA:DNA hybrid.

DSG Coupling Does Not Affect the Recognition of the Substrate by RNase H. The specificity of cleavage by HIV-1 RNase H of both the mimic DNA−RNA:DNA hybrid and its DSG derivative were determined to eliminate the possibility that a structural change on the substrate could alter the recognition by RNase H. The DSG derivative was quenched with glycine prior to the analysis to prevent its cross-linking to the protein. Both hybrids were incubated with RNase H under the conditions used for catalytic activity. A time course analysis of RNase H cleavage is shown in Figure 2. The cleavage pattern of the modified substrate is identical to that observed with the unmodified substrate. Complete cleavage of the modified substrate was observed with higher levels of enzyme, eliminating the possibility that the cleavages detected were due to the presence of a low amount of unmodified substrate (data not shown). These results indicate that the introduction of the cross-linker DSG on the modified base C2dT within the DNA−RNA:DNA hybrid does not affect its recognition by the isolated HIV-1 RNase H domain.

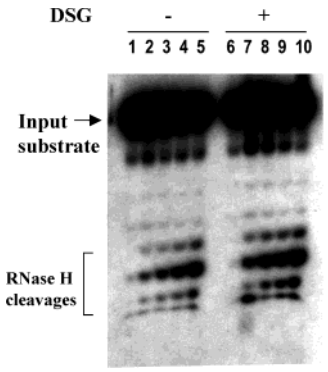

FIGURE 2: tRNA removal activity of NY 427 on the DSG-modified substrate. The tRNA removal activity was determined as described (*9*). After DSG was coupled, the substrate was quenched with glycine and washed by ethanol precipitation. Lanes 1−5, time course analysis of unmodified substrate. Lanes 6−10, DSG-modified substrate. Lanes 1 and 6 ($T = 0$), lanes 2 and 7 ($T = 1$ min), lanes 3 and 8 ($T = 5$ min), lanes 4 and 9 ($T = 10$ min), and lanes 5 and 10 ($T = 20$ min).

Cross-Linking Kinetics. Figure 3 shows the results of a time course analysis of cross-linking in the presence of equimolar levels of RNase H to DNA−RNA:DNA hybrid, labeled on the 5′ end of the modified DNA strand. The products of the reaction were analyzed by SDS−PAGE (Figure 3A). A covalent adduct of approximately 23 kDa is the predominant species detected (Figure 3A). This corresponds with the predicted size of a 1:1 protein/nucleic acid ratio under denaturing conditions, in which the NY427 RNase H contributes 17,050 Da and the DNA contributes 4400 Da. The maximum cross-linking in these conditions was 6.0% after 10 min. Similar results were observed by measuring the retention of the covalent complex on nitrocellulose filters (Figure 3B).

Cross-Linking between RNase H and the DSG-Substrate Derivative Is Dependent on Mn. The isolated RNase H domain (NY427) requires manganese for activity. The effect of varying the Mn^{2+} concentration on the protein−nucleic acid cross-linking was analyzed. The concentration of Mn^{2+} in the reaction was varied between 0 and 30 mM. In the absence of Mn^{2+} no cross-linking was observed (Figure 4A, lane 1). The maximal cross-linking was achieved between 4 and 16 mM Mn^{2+} (Figure 4A, lanes 3−6). At higher concentration of Mn^{2+}, cross-linking decreased. Similar results are observed by filtration on nitrocellulose (Figure 4B). This cross-linking profile parallels the activity profile for NY427 RNaseH, where maximal activity was found between 8 and 16 mM Mn^{2+} (*10*). These results suggest a structural role for Mn^{2+} that contributes to substrate binding (see Discussion).

Cross-Linking Is Competed by Unmodified Substrate. Competition binding experiments were performed to analyze the effects of excess unmodified substrate and nonspecific

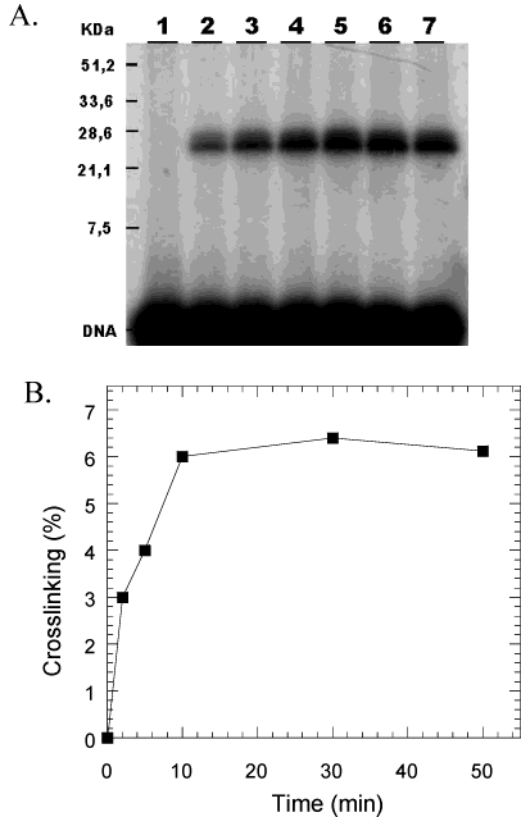

FIGURE 3: Time course of cross-linking. Cross-linking was carried out as described in Materials and Methods. Aliquots of the cross-linking reaction were analyzed by SDS−PAGE (A) and by nitrocellulose binding (B). (A) Lanes 1−7 correspond to $T = 0$, 0.5, 2, 5, 10, 30, and 50 min.

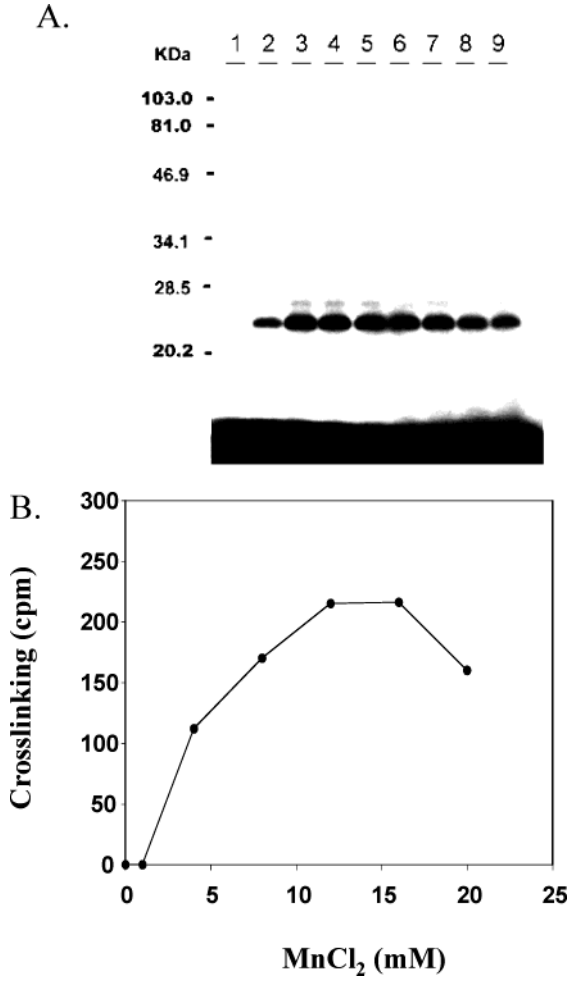

FIGURE 4: Effect of manganese on cross-linking. (A) The concentration of manganese was varied in the cross-linking reaction and the reaction products were analyzed by electrophoresis in 12% acrylamide gels with 0.1% SDS. The gels, after drying, were exposed on X-ray films. Lanes 1−9, the concentration of $MnCl_2$ in the individual lanes: lane 1, 0 mM; lane 2, 1 mM; lane 3, 4 mM; lane 4, 8 mM; lane 5, 12 mM; lane 6, 16 mM; lane 7, 20 mM; lane 8, 25 mM; and lane 9, 30 mM. (B) Titration of Mn^{2+} on cross-linking analyzed by filtration on nitrocellulose filters.

DNA on the efficiency of nucleic acid−protein cross-linking. In these experiments, unmodified substrate was mixed with the C2dT-modified substrate prior to cross-linking. The ratio of unmodified to C2dT-modified substrate was varied to a maximal ratio of 10:1. Unmodified substrate inhibited cross-linking in a concentration-dependent fashion (Figure 5A), as expected from competition for functional binding to RNase H. The unlabeled DNA−RNA:DNA hybrid more efficiently inhibited cross-linking than a double strand DNA (Figure 5B).

The RNA within the Cross-Linked DNA−RNA:DNA Hybrid Is Not Cleaved. The fate of the RNA within the cross-linked complex was determined. The cross-linking reaction was performed utilizing a ^{32}P-end labeled RNA and the RNase H-nucleic acid complex was isolated under native gel electrophoresis conditions. The RNA was released from the complex and analyzed on a 20% PAG containing 7 M urea (Figure 6). The RNA released from the isolated RNase H−nucleic acid complex corresponded with the size of the intact RNA. The tailing of the experimental sample (lane 3) was due to salt in the eluted fraction. These results indicate that the cross-linking occurred prior to RNase H cleavage and that the RNase H is not active within the cross-linked complex.

The Main Cross-Linking Site Is within the RNase H Domain. Cross-linking could occur through the α-amino group at the N-terminus of the protein. The extent of cross-linking to N-terminus was readily analyzed by thrombin cleavage. The NY427 construct contains a His-tag at the N-terminus followed by a thrombin cleavage site. The RNaseH−nucleic acid complex was generated and subjected to thrombin digestion prior to electrophoresis in polyacrylamide gel. Figure 7 shows that removal of the His tag by thrombin causes a slight decrease on the size of the ^{32}P-labeled protein−DNA complex. Cross-linking through the N-terminus would release a low molecular weight product. The maintenance of an approximately 22 kDa labeled species indicates that cross-linking occurred within the RNase H domain.

Peptide Isolation and Sequencing. To characterize the cross-linked peptides, the covalent complex was isolated by electrophoresis in polyacrylamide and digested with trypsin. Samples were analyzed by electrophoresis on 7 M−urea/17% acrylamide gels (Figure 8A). A main peptide−nucleic acid adduct was observed (indicated by the arrow). In addition, some free nucleic acid is observed and another minor band migrating near the free nucleic acid. The DNA-bound peptides were purified by ionic exchange and desalted on C-18 and Sephadex G-25 and sequenced (Keck Facility, Yale University). The N-terminal sequence identified was VVTLTDTTNQ, which matches the sequence Val466−Gln475 of the peptide Val466−Lys517 of HIV-1 RT. On

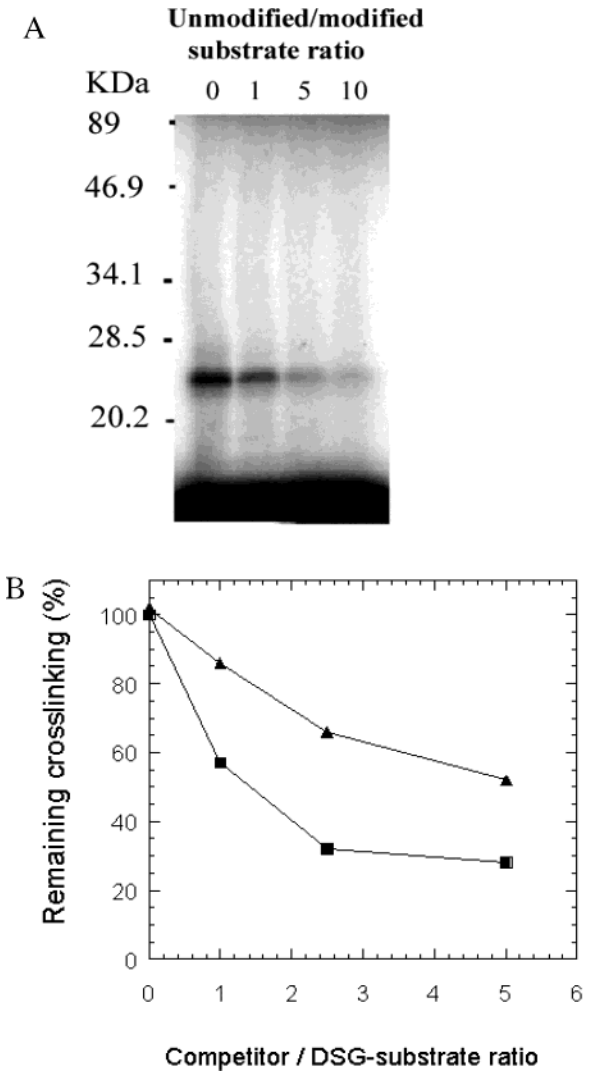

FIGURE 5: Competition of cross-linking by unmodified substrate and nonspecific DNA The cross-linking reaction was carried out as described in methods using 1 pmol of NY427 and 1 pmol of DSG-modified substrate. (A) Unmodified substrate. Products were analyzed by SDS−PAGE. Positions of the protein molecular weight standards are indicated on left. (B) Unmodified substrate (squares) and 21-mer dsDNA (triangles). Cross-linking products were analyzed using a BioRad phosphorimager. Ratios of competitor/DSG-modified substrate are indicated.

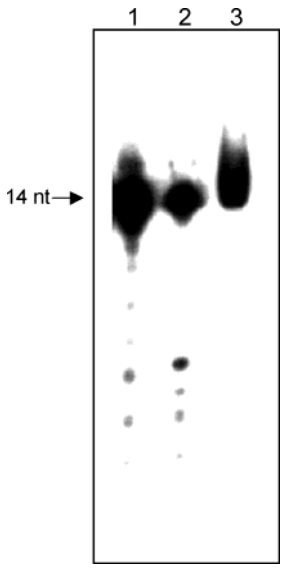

FIGURE 6: Analysis of the RNA in the cross-linked complex. The RNA strand within the DNA−RNA:DNA substrate was ^{32}P-labeled. After cross-linking, the covalent complex was isolated by gel electrophoresis under native conditions and the RNA was eluted and subjected to electrophoresis on a 10% PAG containing 7 M urea. Lane 1, ^{32}P-labeled RNA. Lane 2, unbound substrate. Lane 3, RNA released from the cross-linked complex.

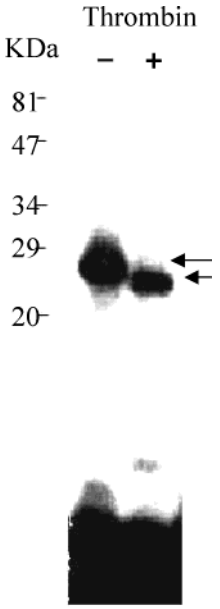

FIGURE 7: Thrombin digestion of the cross-linked complex. A small aliquot of the cross-linking reaction mixture was incubated with thrombin as described in methods, and the products were analyzed by SDS−PAGE. The arrows indicate the mobility of the digested and undigested complexes.

the basis of the fact that the blockage of any lysine by reaction with DSG stops trypsin digestion and that this peptide can be generated by cleavage of lysine residues at both ends, we conclude that the cross-linked amino acid is the internal Lys476. The isolated peptide contains the amino acids Glu 478 and Glu 498, which are part of the catalytic triad. Lys476 was mutated to cysteine, to definitively identify it as the cross-linking amino acid. The NY427 RNase H bearing K476C was expressed, purified, and analyzed for catalytic as well as cross-linking activity. Mutation of Lys476 to cysteine in NY427 resulted in more than 95% decrease of cross-linking (Figure 8C), while the enzyme retained approximately 20% of the activity (Figure 8B). This confirms the role of K476 in cross-linking to the tRNA:DNA mimic substrate.

DISCUSSION

During viral replication, the RNase H domain of HIV-1 RT catalyzes the cleavage of the RNA within RNA:DNA hybrids, both in a polymerase-dependent or -independent fashion. Cleavage occurring at a distance of 18−20 nucleotides behind DNA polymerization is referred to as polymerase-dependent cleavage. However, cleavages of the viral RNA for the initiation of second DNA strand synthesis and for removal of the primer tRNA may occur in the absence of polymerization (polymerase independent).

The protein and substrate requirements for the specific removal of the tRNA primer have been under extensive study. The RNase H cleavage is extremely specific, with the initial cleavage occurring between the penultimate C and the

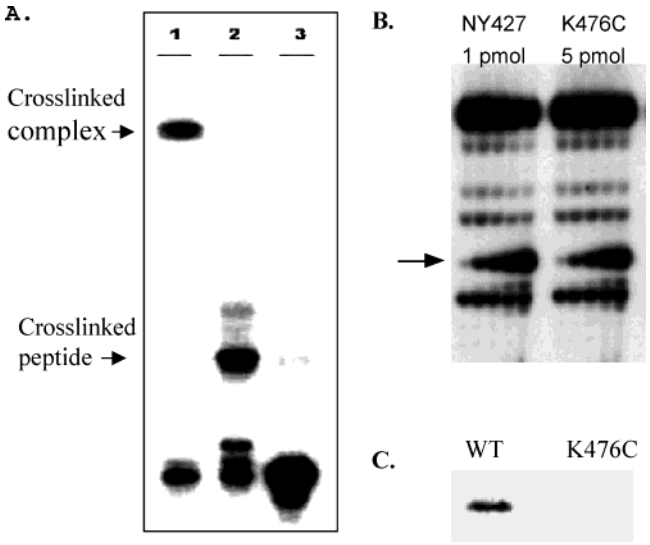

FIGURE 8: (A) Tryptic digestion of the isolated RNase H−nucleic acid complex. The covalent complex was purified by electrophoresis in a 12% acrylamide gel, eluted by diffusion in water, concentrated in a speed vac, and digested with trypsin (3%) in 0.1 M NH_4HCO_3, 2 M urea. Samples were electrophoresed in 17% acrylamide−7 M urea gels. Lane 1, undigested complex, lane 2, digested complex, lane 3, labeled DSG modified substrate. (B and C) Characterization of K476C for RNase H cleavage and cross-linking to the DNA−RNA:DNA hybrid. (B) RNase H activity. Time course analysis of RNA cleavage samples were analyzed at $T = 0$, 2, 5, 15, and 30 min. The arrow marks the released RNA product. (C) Protein−nucleic acid cross-linking. Cross-linking analysis was as described in Materials and Methods. Cross-linked products are shown after electrophoresis on a 15% SDS−PAG. Protein samples are as indicated.

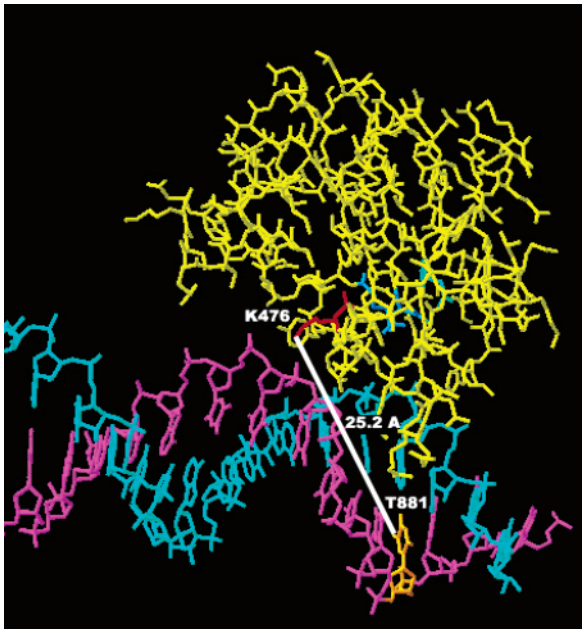

FIGURE 9: Diagram illustrating the positions of Lys476 (red) and the modified base C2dT (orange) in the PPT−RT complex. The distance is measured from the ϵ-amino group of lysine and the 5-methyl group of thymine. The diagram was generated with the program DeepView/Swiss-PDBViewer (GlaxoSmithKline) and PDB ID code 1HYS.

terminal A of the $tRNA^{Lys,3}$ (*15*, *16*). The cleavage at this position results in the transfer of the terminal A of the tRNA to the terminus of the viral DNA and the generation of the two base single-strand tail upon cleavage of the viral DNA by the viral integrase. The specificity of cleavage is identical using either the full-length RT or the isolated RNase H domain (*10*). Optimal site-specific cleavage of the DNA−RNA:DNA mimic requires sequences beyond the first six but less than nine nucleotides of the tRNA. Substitution analysis indicated that positions 4 and 6 from the tRNA 3′ terminus affected the specificity of cleavage (*11*). These studies indicate that the virus has evolved a mechanism to maintain the consistent removal of the tRNA primer and that minimally these determinants localize within the RNase H domain and the 3′ terminus of the tRNA.

In this study, we have utilized this defined, minimal tRNA removal system to identify the peptide regions of RNase H at the nucleic acid binding site. We have extended the system to include a chemical cross-linking agent within the model substrate containing the modified base C2dT on the DNA strand that pairs to the riboadenosine cleaved during removal of the tRNA primer. This modified base was derivatized with DSG for cross-linking directed to amino groups. We found that cross-linking occurred only in the presence of Mn^{2+}, similarly to enzymatic activity as previously reported. This result not only suggests a specific binding but also that Mn^{2+} provides a structural role in NY427.

The cross-linked complex was purified, digested with trypsin, and the DNA-bound peptide was isolated by ion-exchange chromatography. A main cross-linking product was observed in sequencing gels. The N-terminal sequence of the peptide was determined by Edman degradation yielding the sequence VVTLTDTTNQ. According to the specificity of trypsin and the reactivity of the *N*-hydroxysuccinimide ester, the cross-linked lysine would be Lys476. This amino acid forms hydrogen bonds with a phosphate group of the DNA strand of the RT−PPT RNA:DNA hybrid (*4*).

Using the structure of the RT−PPT RNA:DNA hybrid (*4*), several interesting questions arise. The maximum length of the cross-linker used in this work is 15 Å, and it extends from the major groove from position 6 of C2dT. The minimum distance between position 6 of C2dT and the ϵ-amino group of Lys476 in the structure of the complex RT−RNA:DNA would be 25 Å (Figure 9). Furthermore, it is not possible to observe a straight path of the cross-linker. On the basis of the RT−PPT complex, therefore, the distance between Lys476 and the modified base would not accommodate a cross-linking reaction. These results imply that the structure of the RT−PPT RNA:DNA hybrid is distinct from that of the tRNA:DNA sequence that is specifically recognized by the HIV-1 RNase H.

It should be noted that the PPT sequence, within the virus, is resistant to cleavage by RNase H. One possibility is that within the isolated RNase H domain, there is added flexibility and that the Mn^{2+} induced conformation positions the Lys476 in closer proximity with the scissile bond. In addition, the structure of the tRNA:DNA hybrid may be an important contributing factor. Structural studies of related RNA:DNA hybrids indicate that they do not form classic A or B structures, in particular at the RNA−DNA junctions (*17*, *18*). Quite significantly, the sequences that define the recognition of $tRNA^{Lys,3}$ over $tRNA^{Pro}$ lie outside of the region of the scissile bond (*11*). Therefore, the structure and the positioning of the DNA−RNA:DNA mimic for tRNA removal may be quite distinct from that of the PPT.

An alternative possibility is that the binding of the substrate may be in a different orientation. Changing the orientation of the RNA:DNA hybrid in the complex with RT in 180° along the nucleic acid axis and based on the structure of the free hybrid in solution (*9*), the distance between C2dT and Lys476 would be within the range expected for cross-linking. In this model, the cross-linker arm can be easily extended through the major groove to react with Lys476. Definite proof of an alternate orientation will require further investigation using other cross-linking sites. Experiments are currently underway to address this question.

For tRNA removal, the reverse transcription is stopped at the m1A modified base of the tRNA (*19*). The effect of this pausing on the RNase H mode of action is not known; however, it is clear that the RNA within the DNA−RNA:DNA mimic can be specifically cleaved in a polymerase-independent mode using short substrates (*20*). Footprinting studies (*21*) and enzymatic analysis (22) have shown that RT binds to the 5′ end of the RNA strand on a RNA:DNA hybrid. Analysis of the cleavage products indicates that the first cleavage occurs at 18 nucleotides from the 5′ end, and then a second cleavage is made within the first product at eight nucleotides from the 5′ end. Since the polymerase site is located at 18 nucleotides from the RNase H site, it is possible that the first cleavage depends on binding of the polymerase site; however, the secondary cleavage can be produced by RT sliding or rebinding. The cross-linking of K476 to the DNA can occur during a related sliding process, as the RNA remains uncleaved within the covalent complex. This is consistent with the localization of K476 within the primer grip region (4), which acts to guide the RNA template into the RNase H active site. It is important to note that K476A mutation has been extensively studied in vitro and in vivo and resulted in minor (<2-fold) decrease in viral titer (*23*) and minimal perturbation in RNase H function (*24*). This is consistent with the characteristics presented for the K476C mutant, in particular since the chemical modification of the substrate or the mutagenesis of the amino acid did not alter the specificity of the RNase H cleavage. In addition, the lack of cleavage of the substrate eliminates the possibility that the cross-linking occurred during the release of the product from the enzyme. The results from these studies therefore identify a state within the RNase H catalytic reaction in which the positioning of the substrate within the RNase H active site brings the DNA at the scissile bond within 15 Å of Lys476, within the primer grip.

REFERENCES

1. Telesnitsky, A., and Goff, S. P. (1997) in *Retroviruses* (Coffin, J. M., Hughes, S. H., and Varmus, H. E., Eds.) pp 121−160, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
2. Mishima, Y., and Steitz, J. A. (1995) *EMBO J. 14*, 2679−2687.
3. Huang, H., Chopra, R., Verdine, G. L., and Harrison, S. C. (1998) *Science 282*, 1669−1675.
4. Sarafianos, S. F., Das, K., Tantillo, C., Clark, A. D., Jr., Ding, J., Whitcomb, J. M., Boyer, P. L., Hughes, S. H., and Arnold, E. (2001) *EMBO J. 20*, 1449−1461.
5. Han, G. W., Kopka, M. L., Cascio, D., Grzeskowiak, K., and Dickerson, R. E. (1997) *J. Mol. Biol. 269*, 811−826.
6. Coté, M. L., Pflomm, M., and Georgiadis, M. M. (2003) *J. Mol. Biol. 330*, 57−74.
7. Arnott, S., Chandrasekaran, R., Millane,, R. P., and Park, H. S. (1986) *J. Mol Biol. 188*, 631−640.
8. Federoff, O. Y., Salazar, M., and Reid, B. R. (1993) *J. Mol. Biol. 233*, 509−523.
9. Horton, N. C., and Finzel, B. C. (1996) *J. Mol. Biol. 264*, 521−533.
10. Smith, J. S., and Roth., M. J. (1993) *J. Virol. 67*, 4037−4049.
11. Smith, C. S., Leon, O., Smith, J. J., and Roth, M. J. (1998) *J. Virol. 72*, 6805−6812.
12. Leon, O., and Schulman, L. H. (1987) *Biochemistry 26*, 5416−5422.
13. Schulman, L. H., Valenzuela, D., and Pelka, H. (1981) *Biochemistry 20*, 6018−6023.
14. Laemmli, U. K. (1970) *Nature 227*, 680−685.
15. Pullen, K. A., Ishimoto, L. K., and Champoux, J. J. (1992) *J. Virol. 66*, 367−373.
16. Smith, J. S., and Roth, M. J. (1992) *J. Biol. Chem. 267*, 15071−15079.
17. Salazar, M., Champoux, J. J., and Reid, B. R. (1993) *Biochemistry 32*, 739−744.
18. Fedoroff, O. Y., Salazar, M., and Reid, B. R. (1996) *Biochemistry 35*, 11070−11080.
19. Auxilien, S., Keith, G., Le Grice, S. F., and Darlix, J. L. (1999) *J. Biol. Chem. 274*, 4412−4420.
20. Snyder, C. S., and Roth, M. J. (2000) *J. Virol. 74*, 9668−9679.
21. DeStefano, J. J., Cristofaro, J. V., Derebail, S., Bohlayer, W. P., and Fitzgerald-Heath, M. (2001) *J. Biol. Chem. 276*, 32515−32521.
22. Wisniewski, M., Balakrishnan, M., Palaniappan, C., Fay, P. J., and Bambara, R. A. (2000) *J. Biol. Chem. 275*, 37664−37671.
23. Julias, J. G., McWilliams, M. J., Sarafianos, S. G., Arnold, E., and Hughes, S. H. (2002) *Proc. Natl. Acad. Sci. U.S.A. 99*, 9515−9520.
24. Rausch, J. W., Lener, D., Miller, J. T., Julias, J. G., Hughes, S. H., and Le Grice, S. F. (2002) *Biochemistry 41*, 4856−4865.

BI035454Y